

## EVALUACIÓN DE MÉTODOS DE HOMOGENEIZACIÓN DE SERIES CLIMÁTICAS DIARIAS EN EL MARCO DEL PROYECTO INDECIS

J.A. GUIJARRO<sup>1</sup>, N. PÉREZ-ZANÓN<sup>2</sup>, E. AGUILAR<sup>2</sup>, J. SIGRÓ<sup>2</sup>,  
T. CALOIERO<sup>3</sup>, S. CARVALHO<sup>4</sup>, R. COSCARELLI<sup>3</sup>, M. CURLEY<sup>5</sup>,  
E. ENGSTROM<sup>6</sup>, L. LLEDÓ<sup>7</sup>, J. RAMÓN<sup>7</sup>, P. ŠTÉPANEK<sup>8</sup>,  
M.A. VALENTE<sup>4</sup>, G. VAN DER SCHRIER<sup>9</sup>, P. ZAHRADNICEK<sup>8</sup>

<sup>1</sup>State Meteorological Agency, Spain<sup>2</sup>Universitat Rovira i Virgili, Spain; <sup>3</sup>Consiglio Nazionale delle Ricerche, Italy; <sup>4</sup>Universidade de Lisboa, Portugal; <sup>5</sup>Met Eireean, Republic of Ireland; <sup>6</sup>Swedish Meteorological and Hidrological Institute, Sweden; <sup>7</sup>Barcelona Supercomputing Center, Spain; <sup>8</sup>Global Change Research Institute, Czech Republic; <sup>9</sup>Royal Netherlands Meteorological Institute, The Netherlands

[jguijarrop@aemet.es](mailto:jguijarrop@aemet.es)

### RESUMEN

El proyecto INDECIS (*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism*) precisa abordar la homogeneización y control de calidad de series diarias de las variables climáticas esenciales almacenadas en ECAD. Las series de dos regiones diferentes (sur de Suecia y Eslovenia) se analizan para evaluar el tipo, magnitud y frecuencia de las inhomogeneidades a introducir en series de prueba homogéneas generadas a partir del Modelo Climático Regional RACMOv2. Tras aplicarles los métodos de homogeneización disponibles, se compararán sus resultados mediante métricas de bondad de ajuste y se seleccionarán las mejores metodologías para depurar las series almacenadas en ECAD, con objeto de calcular índices climáticos relevantes con los que evaluar el impacto del cambio climático en sectores económicos prioritarios.

**Palabras clave:** homogeneización, series climáticas diarias, proyecto INDECIS.

### ABSTRACT

The INDECIS project (*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism*) needs to address the homogenization and quality control of daily series of the essential climatic variables stored in ECAD. The series of two different regions (southern Sweden and Slovenia) are analyzed to evaluate the type, magnitude and frequency of the inhomogeneities to be introduced in homogeneous test series generated from the Regional Climate Model RACMOv2. After applying the available homogenization methods, their results will be compared by means of goodness of fit metrics and the best methodologies will be selected to debug the series stored in ECAD, in order to calculate relevant climatic indices with which to evaluate the impact of climate change in priority economic sectors.

**Key words:** homogenization, daily climate series, INDECIS project.

## 1. INTRODUCCIÓN

El estudio de la variabilidad climática precisa de series observacionales de alta calidad, para lo que se desarrollaron a lo largo del tiempo distintos métodos de detección y corrección de datos y saltos anómalos producidos por alteraciones de las condiciones de observación (y por tanto ajenos a las variaciones climáticas que se quieren analizar). Paralelamente a ello surgió el interés por comparar estas metodologías, y bastantes autores realizaron experimentos con series sintéticas. La Acción COST ES0601 (“HOME”, 2006-2011), financiada por la *European Science Foundation*, marcó un hito en estos esfuerzos al convocar a homogeneizadores y desarrolladores de todo el mundo para aplicar sus métodos a un extenso conjunto de series de temperaturas y precipitaciones mensuales con diferentes tipos de inhomogeneidades conocidas solo por un investigador de control (Venema *et al.*, 2012).

Las provechosas experiencias de “HOME” alentaron mejoras en los métodos existentes y la aparición de otros nuevos. Además, diferentes variables climatológicas medidas en distintos climas con variadas combinaciones de densidad de estaciones y complejidad de la orografía hacen difícil generalizar los resultados de las comparaciones. Ante la dificultad de repetir un proyecto tan costoso como “HOME”, el proyecto español MULTITEST (Guijarro *et al.*, 2016 y 2017) se centró en aquellos métodos que pudieran aplicarse en modo totalmente automático, para poder realizar centenares de pruebas diversas con series mensuales sintéticas de precipitación y temperatura simulando distintos climas.

La mayor parte de estos esfuerzos se centraron en la homogeneización de series mensuales de temperatura y precipitación, pero actualmente la atención se dirige a la corrección de series diarias, necesarias para el estudio de la variabilidad y tendencias de los fenómenos extremos. Tras un primer intento de comparación al final de la Acción COST “HOME”, Killick (2016) evaluó varias homogeneizaciones de series diarias de temperatura simulando cuatro regiones de EEUU, pero con un reducido número de participantes.

Actualmente, el proyecto europeo INDECIS (*Integrated approach for the development across Europe of user oriented climate indicators for GFCS high-priority sectors: agriculture, disaster risk reduction, energy, health, water and tourism*), liderado por la Universidad Rovira i Virgili (Tarragona) y enfocado al desarrollo de indicadores climáticos orientados a usuarios de sectores económicos prioritarios, precisa abordar la homogeneización y control de calidad de las series diarias de las variables climáticas esenciales almacenadas en el European Climate Assessment and Dataset (ECAD: Klein Tank *et al.*, 2002), con objeto de que posteriormente se puedan calcular los índices climáticos relevantes para evaluar el impacto del cambio climático. En los siguientes apartados se presentará y discutirá la metodología proyectada y los resultados esperados.



## 2. MÉTODOS

Para la comparación de los métodos de homogeneización y control de calidad de las series diarias son necesarias dos cosas: Una base de datos con series diarias homogéneas de las variables climáticas que se vayan a estudiar, y un catálogo de alteraciones aplicables a esas series para hacerlas inhomogéneas replicando la casuística de las series reales. Después bastará generar las series de prueba y homogeneizarlas con las metodologías disponibles, comparar sus resultados con las series originales, y calcular métricas para valorarlos y compararlos entre sí.

Las variables climáticas almacenadas en ECAD con resolución diaria y las unidades en que están expresadas son:

- CC: Nubosidad (octavos de cielo cubierto)
- DD: Dirección del viento (°)
- FG: Velocidad del viento (0.1 m/s)
- FX: Racha máxima del viento (0.1 m/s)
- HU: Humedad relativa (%)
- PP: Presión atmosférica reducida al nivel del mar (0.1 hPa)
- RR: Precipitación (0.1 mm)
- SD: Espesor de la nieve (cm)
- SS: Horas de insolación (horas)
- TG: Temperatura media (0.1 °C)
- TN: Temperatura mínima (0.1 °C)
- TX: Temperatura máxima (0.1 °C)

### 2.1. Base de datos homogénea

Dado que no se puede garantizar que las series reales sean homogéneas, para disponer de una base de datos homogénea se ha recurrido tradicionalmente a generar series sintéticas procurando que tengan las mismas propiedades estadísticas de las series reales de una determinada zona (tal como se hizo en los proyectos HOME y MULTITEST ya citados). Otra posibilidad sería usar una base de datos homogeneizada, pero podrían quedar inhomogeneidades residuales, además de no quedar asegurarse que el método usado en la homogeneización no vaya a estar favorecido en la posterior comparación. Por todo ello aquí se van a usar series obtenidas a partir del Modelo Climático Regional (MCR) RACMOv2 (KNMI), ligado al Modelo Global MOHC-HadGEM2-ES (Collins et al., 2008), elegido por su alta resolución (0,11°) y por ofrecer salidas de todas las variables climáticas almacenadas en ECAD.

Para generar los datos de prueba se han seleccionado dos áreas con alta densidad de series observadas y climas diferentes: Frio (Sur de Suecia) y mediterráneo húmedo (Eslovenia), y se ha procurado que las series del MCR tengan características similares a las de las observaciones de su propia área.

### 2.2. Inhomogeneidades a introducir

Para ambas áreas experimentales y para cada una de las variables a estudiar, se generan seis redes, combinando correlaciones buenas y medianas con tres problemáticas de dificultad creciente: a) solo saltos en la media; b) saltos y lagunas de datos; y c) saltos, lagunas y valores anómalos.

Los saltos en la media se generan aleatoriamente a partir de una distribución de Poisson con un promedio de 4, y su magnitud se modela de acuerdo a las características deducidas del estudio de series paralelas obtenidas del proyecto ISTI/POST ([http://www.surfacetemperatures.org/databank/parallel\\_measurements](http://www.surfacetemperatures.org/databank/parallel_measurements)), cuyas diferencias sirven de modelo para simular el impacto que tendrían los cambios de emplazamiento de las estaciones en sus series observadas.

### **2.3. Métodos a probar y procedimiento a seguir**

Los programas de homogeneización a probar serán, por orden alfabético, ACMANT (en sus versiones para temperatura y precipitación), Climatol, HOMER, MASH y SplidHOME. Las redes inhomogéneas generadas por el procedimiento explicado anteriormente se suministrarán a los participantes especializados en cada método para que los apliquen y devuelvan resultados homogeneizados, que se compararán con las soluciones (series homogéneas originales), calculando el error típico y comparando sus tendencias, promedios, desviaciones típicas, percentiles extremos, etc.

Paralelamente, la infraestructura de programación que se usó en el proyecto MULTITEST, adaptada para su utilización con las series diarias del banco de pruebas de INDECIS, va a permitir aplicar los métodos a un gran número de muestras aleatorias de series de cada una de las variables climáticas estudiadas, con inhomogeneidades introducidas de forma similar. Esto solo se puede llevar a cabo con procedimientos totalmente automatizados, como es el caso de ACMANT y Climatol, que por otra parte fueron los que dieron mejores resultados en la comparación realizada por Killick (2016). Pero cualquier otro método puede también participar en estas pruebas intensivas con tal de que algún especialista proporcione las rutinas de programación necesarias para su aplicación.

## **3. RESULTADOS**

### **3.1. Idoneidad de las series del Modelo Climático Regional**

Para contrastar la idoneidad de las series del MCR se han comparado con series observadas de su misma área geográfica, observando que sus características son similares. En la figura 1 se muestra un ejemplo de una serie de un punto de rejilla del sur de Suecia junto a la serie observacional más próxima, presentando una gran similitud, confirmada con las distribuciones acumuladas empíricas de sus valores (no mostradas aquí).

Otro aspecto a tener en cuenta para que las series del MCR sean aptas para servir como banco de datos es que las correlaciones entre las series de la muestra sean similares a las de las series observadas. La figura 2 muestra los correlogramas de las series del MCR del sur de Suecia para las 12 variables almacenadas en ECAD, mientras que en la figura 3 se presentan algunos correlogramas de las series observadas en la misma zona. Se puede ver que existe cierta similitud, pero los correlogramas de las series reales presentan mayor dispersión y degradación de la correlación al aumentar la distancia entre las estaciones de medida, que en parte puede estar causada por las inhomogeneidades contenidas en ellas. En esta misma figura se ilustran tres posibles criterios para hacer que las correlaciones del banco de

datos de prueba sea asimilable a las que se tienen en las series reales: 1) la correlación media a 200 km de distancia; 2) la correlación máxima a 400 km; y 3) la mediana de las correlaciones.

Para hacer más similares los correlogramas de las series de prueba se puede variar la distancia entre los puntos de rejilla del MCR y también añadir ruido estadístico a las series

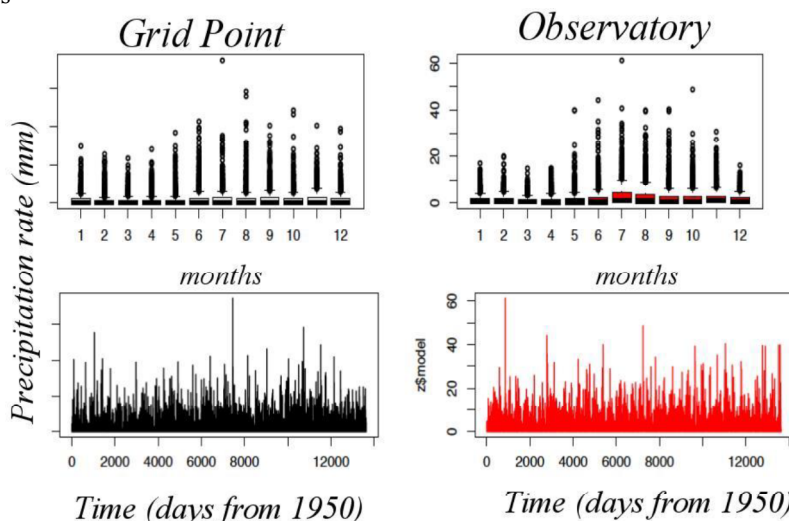


Fig. 1: Serie de precipitación obtenida del MCR (izquierda) y la serie observada más próxima (derecha). Arriba se muestran sus diagramas de caja mensuales, y debajo los valores de ambas series diarias.

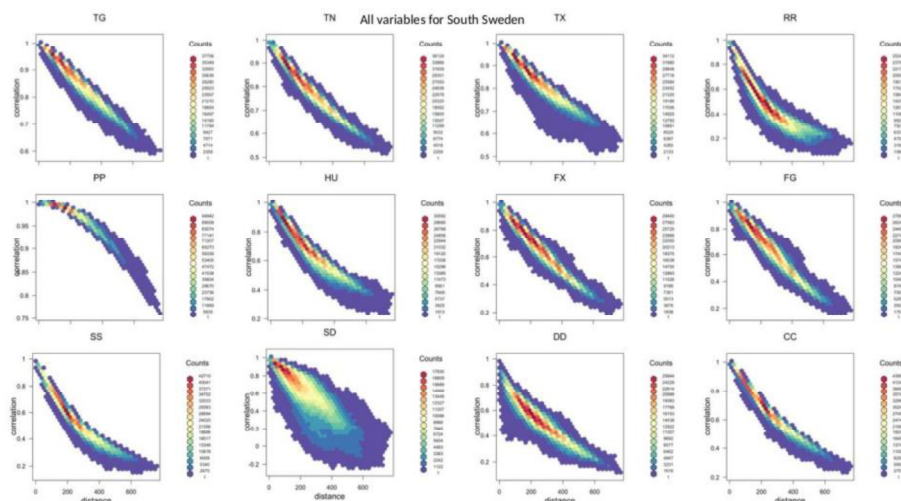


Fig. 2: Correlogramas de las series del MCR del sur de Suecia para las 12 variables almacenadas en ECAD.

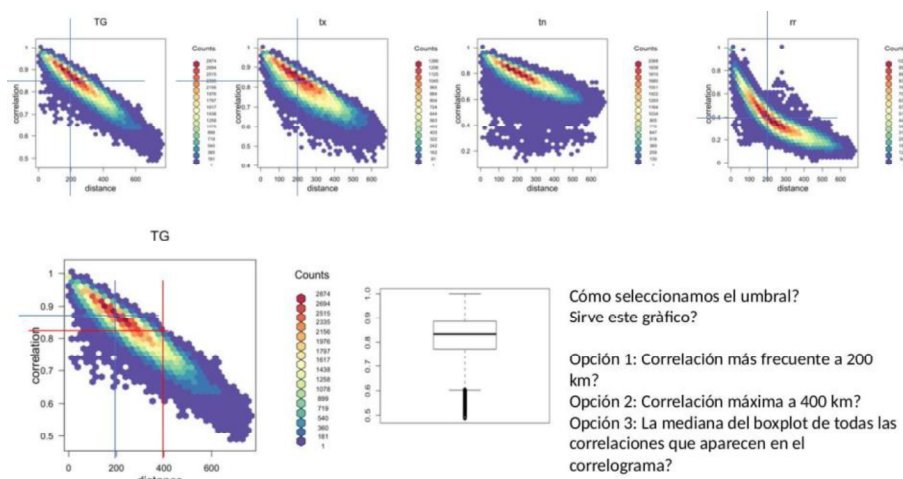


Fig. 3: Correlogramas de las series observadas de temperaturas medias, extremas y precipitación en el sur de Suecia, y posibles criterios para seleccionar un umbral de equiparación con las series del MCR.

### 3.2. Diferencias entre las series paralelas

Las magnitudes de los sesgos a introducir en las series del banco de datos de prueba para inhomogeneizarlas se pueden extraer de la comparación entre las series paralelas de ambas zonas geográficas. La figura 4 muestra las diferencias entre los deciles de cada pareja de series paralelas del sur de Suecia para ocho de las variables a estudiar. También se han evaluado estas diferencias por estaciones del año, puesto que es bastante habitual que los sesgos introducidos por cambios en las condiciones de observación tengan distinta magnitud a lo largo del año.

Alternativamente se ha elaborado un catálogo de diferencias medias diarias, de donde poder elegir las al azar para aplicarlas a segmentos de las series del banco de datos de prueba, consiguiendo de este modo replicar el efecto que tendría un cambio de emplazamiento con sesgos potencialmente distintos para cada día del año.

## 4. DISCUSIÓN

Aunque es un proyecto todavía en desarrollo, promete ser un gran paso adelante en la homogeneización de series diarias de diversas variables climáticas, algunas no abordadas anteriormente.

El uso de series procedentes de las salidas de un Modelo Climático Regional permite eliminar las incertidumbres inherentes al empleo de series observadas, y se han mostrado como una buena alternativa al empleo de series sintéticas, a las que es más difícil incorporar las propiedades estadísticas de las series reales.

Es cierto que la variabilidad espacial de las series del MCR es menor que las de las series reales, dando lugar a mayores y menos dispersas correlaciones entre ellas, pero esto se puede corregir aumentando la distancia entre los puntos de rejilla seleccionados y añadiendo ruido blanco en caso necesario.

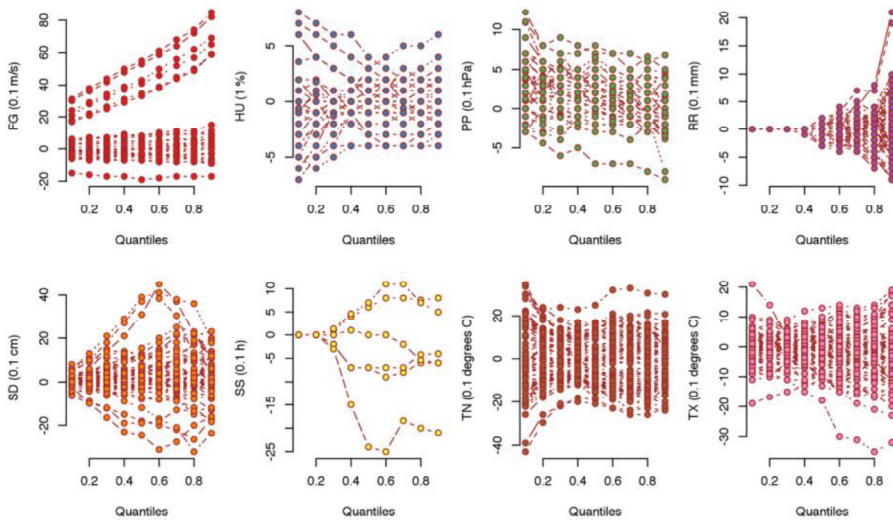


Fig. 4: Diferencias de los deciles entre cada par de series paralelas en el sur de Suecia para ocho de las variables a estudiar.

Es de desear que se puedan aplicar todos los métodos de homogeneización diaria disponibles actualmente, con objeto de poder comparar sus fortalezas y debilidades. La aplicación de los métodos seleccionados a la homogeneización de las series de ECAD permitirá estudiar la variabilidad y tendencias de índices climáticos clave para sectores como la agricultura, energía, sanidad, turismo y recursos hídricos, así como para la reducción del riesgo de desastres.

## AGRADECIMIENTOS

El proyecto INDECIS es parte de ERA4CS, un ERA-NET iniciado por JPI Climate, financiado por FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) y co-financiado por la Unión Europea (Grant 690462).

## REFERENCIAS

- Collins, W.J., Bellouin N., Doutriaux-Boucher M., Gedney N., Hinton T., Jones C. D., Liddicoat S., Martin G., O'Connor F., Rae J., Senior C., Totterdell I., Woodward S., Reichler T., Kim J. (2008). *Evaluation of the HadGEM2 model*. MetOffice Hadley Centre Technical Note 74, 47 pp.
- Guijarro, J.A., Domonkos, P., López, J.A., Aguilar, E. y Brunet, M. (2016). Comparación de métodos de homogeneización de series: primeros resultados del proyecto MULTITEST. En Olcina, J., Rico, A.M., Moltó, E. (eds.): *Clima, sociedad, riesgos y ordenación del territorio*. Universidad de Alicante (Spain), Asociación Española de Climatología, ISBN 978-84-16724-19-2, pp. 131-140.

- Guijarro, J.A., López, J.A., Aguilar, E., Domonkos, P., Venema V.K.C., Sigró J. y Brunet, M. (2017). Comparison of homogenization packages applied to monthly series of temperature and precipitation: The MULTITEST project. 9<sup>th</sup> Seminar for homogenization and quality control in climatological databases and 4<sup>th</sup> conference on spatial interpolation techniques in climatology and meteorology (Budapest, April 3-7), WMO WCDMP-No.85, pp. 46-62.
- Killick, R.E. (2016). *Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data*. PhD Thesis, University of Exeter, 249 pp.
- Klein Tank, A.M.G. and Coauthors (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. of Climatol.*, 22, 1441-1453.
- Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P. And Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Clim. Past*, 8, 89-115.